Received: 6 July 2022 | Revised: 15 November 2022 | Accepted: 5 April 2023

Healthcare Technology Letters

DOI: 10.1049/htl2.12045

The Institution of Engineering and Technology WILEY

LETTER

Automatic cardiac arrhythmias classification using CNN and attention-based RNN network

Jie Sun

School of Cyber Science and Engineering, Ningbo University of Technology, Ningbo, Zhejiang, China

Correspondence
Jie Sun, School of Cyber Science and Engineering, Ningbo University of Technology, No. 201, Fenghua Road, Jiangbei District, Ningbo, Zhejiang, China.
Email: sunjie@nbut.edu.cn

Funding information
This research was funded by Scientific Research Fund of Zhejiang Provincial Education Department (grant number Y201839519), Ningbo Natural Science Foundation (grant number 2019A610087].

Abstract

Cardiac disease has become a severe threat to public health according to the government report. In China, there are 0.29 billion cardiac patients and early diagnosis will greatly reduce mortality and improve life quality. Electrocardiogram (ECG) signal is a priority tool in the diagnosis of heart diseases because it is non-invasive and easily available with a simple diagnostic tool of low cost. The paper proposes an automatic classification model by combing convolutional neural network (CNN) and recurrent neural network (RNN) to distinguish different types of cardiac arrhythmias. Morphology features of the raw ECG signals are extracted by CNN blocks and fed into a bidirectional gated recurrent unit (GRU) network. Attention mechanism is used to highlight specific features of the input sequence and contribute to the performance improvement of classification. The model is evaluated with two datasets considering the class imbalance problem constructed with records from MIT-BIH arrhythmia database and China Physiological Signal Challenge 2018 database. Experimental results show that this model achieves good performance with an average F1 score of 0.9110 on public dataset and 0.9082 on subject-specific dataset, which may have potential practical applications.

1 | INTRODUCTION

Cardiac disease has a death rate of 32% more than cancer and other disease. In China, there are 0.29 billion cardiac patients with a rising prevalence [1]. It is important to detect and diagnose early to reduce mortality and improve life quality. Electrocardiography (ECG) is a non-invasive tool for the diagnosis of cardiac abnormalities. Normal heartbeat usually consists of P, Q, R, S and T waves [2]. Different arrhythmias show specific differences in these five waves and the durations between the waves. The unique local morphology and overall trends can be observed in an ECG record and thus the cardiac abnormalities can be recognized by their characteristics. For examples, the T wave of ventricular ectopic beat was significantly higher than that of non-ectopic beat. While long duration between S and T wave and morphological changes of T wave are several diagnostic indexes of myocardial infarction [3].

Deep learning technology provides a new and effective example for making clinical decision-making from pathophysiologic data [4]. Many researches have focused on the topic to explore potential risks of heart attack and made great progress trying to achieve high performance [5]. Although some works have achieved better performance than a human specialist, there are still unsolved challenges which limit the large-scale promotion of automatic classification of arrhythmias:

1. Many cardiac diseases are chronic and change gradually over time. There is a long-term correlation between disease stage and historical medical intervention and treatment. In order to make a correct medical diagnosis, historical information must be taken into account. An automatic diagnosis model will be more precise if the time-dependency progress of disease symptoms is modelled.
2. The morphological characteristics of ECG signals are significantly different among subjects, and strongly depend on their physical condition. For example, a healthy athlete as regular heartbeat frequency of about 60 beats per minute, which is usually considered as sinus bradycardia of ordinary people. In this case, a normal heart condition for the athlete may be incorrectly identified as abnormal. While for some

physically weak people, the normal signals may be regarded as noise. The high variability between patients and the internal variability of heart rate of the same patient will affect the accuracy of abnormality recognition, and the recognition model is easy to have the problem of over-fitting or underfitting. Human heterogeneity should be paid more attention when designing a universal classification model.

3. The performance of a neural network is highly dependent on the training dataset. If the dataset is imbalanced with 98% of normal events and 2% of abnormal events, the model will achieve accuracy of 98% even it classifies all the inputs into normal events while treats the abnormal data as noise and eliminates it. But these 2% data are just the distinguishing symptoms of ECG abnormality.

To address the time-dependency problem of ECG classification task, this paper proposes a model integrating convolutional neural network (CNN) and recurrent neural network (RNN). The convolutional neural network is used to learn the morphological characteristics of ECG raw signal, and the GRU network is used to extract the long-term correlation between the sequence features. To address the human heterogeneity problem, we conduct patient-specific training method. The training is composed of two phases: the first phase is to produce a generalized model with public datasets, and the second phase is to fine-tune the general model to fit the specific subject. To address the data imbalance problem, the model is trained with constructed datasets in which the records numbers of each class are complemented to be equal.

The remainder of the paper is organized as follows. Section 2 introduces related work. The architecture of the proposed work is introduced in detail in Section 3. Section 4 describes the preparation of datasets. Section 5 gives experimental setup and performance evaluation. Finally, a conclusion is presented in Section 6.

2 | RELATED WORK

The abnormality of ECG can be observed from the morphological and temporal variability between cardiac cycles. In order to explore the variabilities, a model should have the capability to extract the local feature of the cardiac cycles and analyse sequential data. In the field of deep learning, there are two widely used networks: convolutional neural network that focuses on feature extraction and recurrent neural network that focuses on sequence analysis.

2.1 | CNN model and its application in ECG diagnosis

CNN is a special kind of neural network characterized by convolutional operators, which makes it perform well in image processing field. Deep CNN has multiple non-linear hidden layers capable of learning the complex relationship between input and output. In recent years, CNN model has been applied in the diagnosis of different cardiac diseases, such as cardiac arrhythmia, myocardial infarction, and heart failure. In the work of Hannun and co-workers [6], a model consisting of 16 convention blocks with residual connections is proposed to detect 12 kinds of heart arrhythmias from the original single lead ECG input, which achieves an F1 average score of 0.837 better than the average of cardiologists (0.780). In the work of Baloglu [7], a deep learning model of 10 layers is proposed for diagnosis of 10 types of myocardial infarction based on 12 lead ECG signals. For congestive heart failure diagnosis, Acharya and his team developed a 11-layer deep convolutional neural network model with 2-s of ECG signals as input, and attained a diagnostic accuracy of 98.97% [8].

2.2 | RNN model and its application in ECG diagnosis

RNN model is suitable for processing temporal and sequential data with the capability of memory of historic information and is widely used in time series processing. However, the original recurrent neural network suffers from gradient vanishing or gradient exploding. Long short-term memory (LSTM) model is a variant of RNN and solves the problems effectively. It introduces gating mechanism including input gate, output gate and forgetting gate to control the pass route of information.

RNNs are natural to model time-dependent correlation of the ECG data, but need to be fed with preprocessed features, including time domain and frequency domain features. Time domain features involve the time length of ECG waves and intervals and associated further processing, such as RR interval and its standard deviation. Maknickas proposed a three-layer LSTM network on pre-computed features including wave width and amplitude and QRS interval length to distinguish Atrial fibrillation from normal signal. The model achieved an average F1 score of 0.78 with 1791 parameters [9].

Frequency domain features involve features extracted with signal processing techniques such as wavelet transformation or Fourier Transform. Chang proposed an LSTM model of 30 hidden units to extract the long-term and short-term characteristics of atrial fibrillation (AF) with the input of transformed spectrograms from 2-lead ECG signal [10]. The spectrogram is achieved by the Short-Term Fourier Transform (STFT) of data on a sliding window basis. The detection accuracy of the model was 98.3%. In the work of Saadatnejad [11], the ECG feature and wavelet feature of ECG signal are extracted and then the features are fed into two LSTM models to classify the record. Sawant proposed a multilabel classification model based on gated recurrent unit (GRU) with time-frequency features extracted by Fourier Bessel Expansion and scattering transform [12].

Instead of hand-crafted features, Yildirim applied an LSTM model with the features coded using conventional autoencoders (CAE) to automatically classify five types of arrhythmias, and achieved a testing accuracy of 99.11% with a reduction of training time from 4.5 to 0.6 h [13].

2.3 | CRNN model and its application in ECG diagnosis

CNN model has shown its powerful and feasible capability in feature extraction, but it is limited in that all input signals should be segmented into fixed size, which will lead to the loss of temporal feature of continuous signal. RNN model is suitable for sequences of variable length, while learning local feature of input signal is a challenging task for it. In view of the distinctive features, it is inspired to take advantage of the combination of the two models, that is, convolutional recurrent neural network (CRNN).

Lui developed a myocardial infarction (MI) classifier which combines convolutional neural network and recurrent neural network [14]. It uses single lead ECG signal acquired form wearable ECG equipment to distinguish 15 kinds of classification. It is found that the sensitivity is improved by 28.0% after adding a recurrent layer compared with the original convolution neural network.

Much effort has been devoted to demonstrate the performance of the integrated structure. To detect diabetes by heart rate, several experiments are conducted in different network architecture [15]. The accuracy of CNN is 93.6%, while the maximum accuracy of CNN-LSTM combination is 95.1%. Three deep learning networks are proposed to classify ECG signal: CNN, LSTM, convolutional LSTM (CLSTM) [16]. The performance of the models was verified using multiple public arrhythmia database. Among the three models, CLSTM model shows the overall superior performance with an accuracy of 97.6%.

Attention mechanisms are a well-known technique in computer vision [17] and natural language processing [18]. Attention is dynamically selected by adaptive weighting according to the importance of the input. The essence of attention is to locate interested information and suppress useless information. The combination of attention layer with CNN and RNN model has been proposed. Sigurthorsdottir et al. proposed a convolutional recurrent neural network. The blocked convolutional layers extracted features, and a bi-directional gated recurrent unit (GRU) layer and an attention layer is applied to aggregate these features into a single feature vector which is used to classification [19]. Qiao developed a model composed of CNN and Bi-LSTM with multilevel attention to find the abnormal variation in beat-, rhythm- and frequency-level [20].

Although CRNN has shown remarkable progress in ECG diagnosis, there are still much potential needing to be exploited. In our method, we augment the CNN module and RNN module with different attention mechanism to calibration the feature learning, and apply fine-tuning technique to achieve the subject-specific classification model. Besides, we re-construct the training dataset to avoid class-imbalance problem.

3 | METHODS

In this work, we present an automatic classification model combining both CNN and RNN for detection of arrhythmias from ECG signals. Figure 1 illustrates the proposed network architecture. The CNN module is applied to extract morphology features of the ECG signal and the gated recurrent unit (GRU) module is applied to model the long-term temporal dependency of the features. In this study, the network accepts heartbeats from raw ECG signal sampled at 360 Hz as input (described in Section 4.1), and outputs a prediction of the arrhythmia type to which the heartbeat belongs. The arrhythmia types used are normal sinus rhythm (NSR), left bundle branch block (LBBB), right bundle branch block (RBBB), atrial premature beats (APB) and premature ventricular contraction (PVC).

3.1 | CNN module with attention

The architecture of CNN module is shown in Figure 2a. The first block consists of a 1D convolutional (1D Conv) layer, a batch normalization (BN) layer and a rectified linear unit (ReLU) layer. The next four blocks have the same structure as a BN layer followed by a ReLU layer, a dropout layer, and a convolutional layer. Each block will start with a BN layer except for the first one which has made the input normalized in the preprocessing phase.

Batch normalization is the operation of transforming the dataset into having zero mean and unit variances to minimize the impact of internal covariate shift 22, which is the phenomena that the input distribution of each layer will change with the parameters of the previous layer in training phase. BN transform can be added to a network to manipulate any activation and enables higher learning rate.

ReLU layer will introduce non-linearity into the model, accelerate convergence speed and improve accuracy. Dropout is a regularization technique to discard units randomly. It will prevent the network from forming complex cooperative adaptation and significantly reduces over fitting and improves accuracy. We apply dropout with a probability of 0.2. The convolution layer contains a 1D convolution layer (each have 32 kernels of size 5). Max pooling is an operation that compute the max value of a particular feature, which reduces the dimensions of the output features significantly while enables translation invariant of the features. We use max pooling of size 5 and stride 2 in all pooling layers to reduce the number of parameters and computation.

The SE (squeeze-and-excitation) module is applied to refine the channel-wise feature maps [21]. SE module consists of a global average pooling (GAP) layer and two fully connection (FC) layers, each with different activation functions. Given the input feature vector as f_{in}, the GAP layer will squeeze global spatial information into a channel descriptor to capture channel-wise dependencies. The SE module will produce a scalar s to represent the importance of the channel as shown in Equation (1), where δ refers to the ReLU function and σ refers to the Sigmoid function. The refined feature vector is shown in Equation (2), where $s{\cdot}f_{in}$ refers to the channel-wise multiplication between the feature vector and the scalar s.

$$s = \sigma\left(W_2\delta\left(W_1 GAP\left(X\right)\right)\right) \tag{1}$$

$$f_{out} = f_{in} + s \cdot f_{in} \tag{2}$$

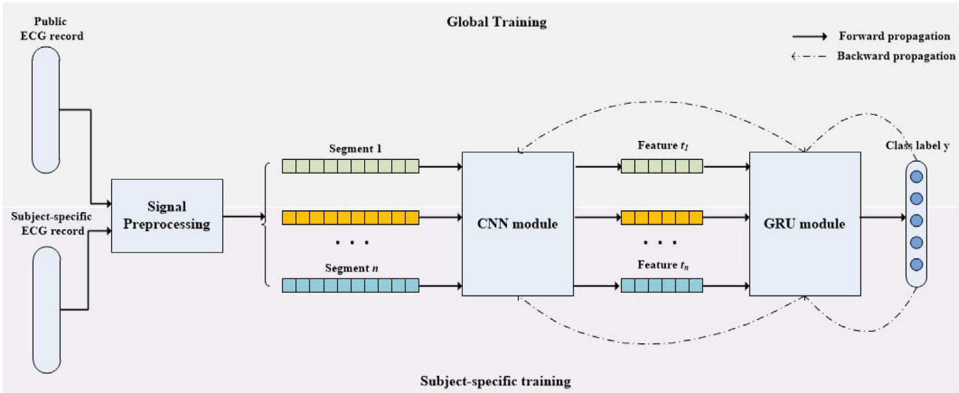

FIGURE 1 The block diagram of our proposed model. The neural network consisted of a CNN module and a GRU module. The training of the model includes global training and subject-specific training.

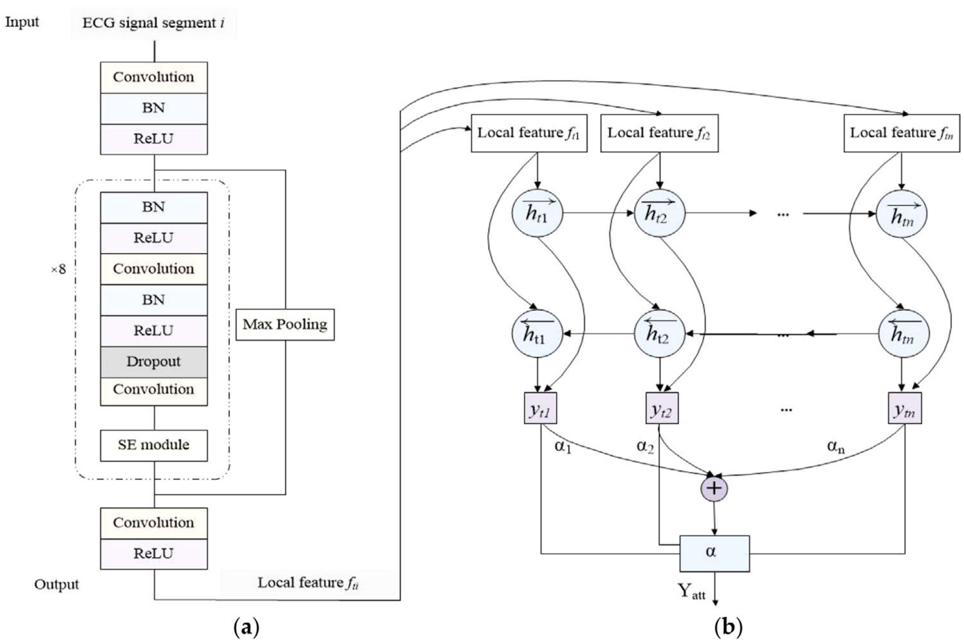

FIGURE 2 (a) The architecture of CNN module. (b) The architecture of GRU module. The output of CNN module of different timestamp is fed into the RNN module for further processing.

3.2 | Attention-based bidirectional GRU

GRU is applied to extract the long-term temporal features along the time sequence to infer the trend of the signal. We apply a bidirectional GRU model to capture the forward and backward time dependency. As shown in Figure 2b, the model is composed of a forward layer and a backward layer. Each layer has two units for different direction. At each time step t, the latent feature vector f_t of an ECG segment is fed into the GRU module. The number of the units is the length of the considered time steps of an ECG record, and set to 32 after several testing of varied setting.

In each GRU unit, the hidden state $\overrightarrow{h_t}$ of time step t in the forward layer and $\overleftarrow{h_t}$ in the backward layer are determined by the current input and the hidden state of the previous time step as denoted in Equations (3) and (4), where $\vec{W}$ and $\vec{V}$ denoting the weight and $\vec{b}$ denoting the bias of the forward layer, and $\overleftarrow{W}$, $\overleftarrow{V}$ and $\overleftarrow{b}$ are that of the backward layer. They are trainable parameters.

The annotation vector of each ECG segment is calculated by summarizing its hidden outputs from both directions as $[\overrightarrow{h_t}; \overleftarrow{h_t}]$ and the output of the unit y_t are computed as Equation (5), where f_t represents the input of GRU in timestep t. The output will be constructed into a matrix Y of size of $N \times T$, where T is the length of the input sequence and N is the size of vector y_t, and fed into the attention layer.

$$\overrightarrow{h_t} = \tanh\left(\vec{W} f_t + \vec{V}\,\overrightarrow{h_{t-1}} + \vec{b}\right) \tag{3}$$

$$\overleftarrow{h_t} = \tanh\left(\overleftarrow{W} f_t + \overleftarrow{V}\,\overleftarrow{h_{t-1}} + \overleftarrow{b}\right) \tag{4}$$

$$y_t = \tanh\left(U\left[\overrightarrow{h_t};\overleftarrow{h_t}\right] + b_y\right) \tag{5}$$

Soft attention mechanism is applied to emphasize information specific to the subject heterogeneity, which is represented with an attention weight vector α denoted in Equation (6). The output from the attention layer is denoted as Equation (7), where the superscript T denotes the transpose of a matrix.

$$\alpha = softmax\left(w_{att}^{T} Y\right) \tag{6}$$

$$Y_{att} = Y\alpha^{T} \tag{7}$$

3.3 | Classification layer

The attention layer is then followed by a fully connected layer, the output of which is fed into a SoftMax layer. The output from the SoftMax layer corresponded to the probability distribution of arrhythmia in the input ECG segment.

3.4 | Subject-specific training

The model is trained by two steps: global training and subject-specific training as shown in Figure 1. The global training uses public dataset to get a generalized model. The subject-specific training uses patient-specific dataset to fine-tune the generalized model and get a subject-specific model. Fine-tuning technique is applied when a pretrained DNN to be reused for new task. The parameters of the model can be adjusted to find a new minimized loss while the structure of the model remains intact. The subject-specific training is performed forward propagation to calculate the loss of the objective function, and then perform back propagation to calculate the gradient of the filter (parameter) using the chain rule. Finally, the parameters are updated by batch SGD algorithm. Forward propagation and backward propagation are performed iteratively until the loss converges.

In implementation, the first five CNN modules are frozen in subject-specific training to avoid overfitting and fasten the training speed. A lower learning rate is adopted to promote the convergence of the network. Therefore, the model can complete training with only dozens of back propagation stages, making the classification task as quickly as possible. This makes the subject-specific model for real-time ECG monitoring applicable.

4 | DATASETS

In this work, the public datasets are obtained from two publicly available database. MIT-BIH Arrhythmia database (MITDB) [22] consists of 48 records obtained from 47 subjects sampled at 360 Hz. Each record contains two leads, that is, modified limb lead II obtained by placing the electrodes on the chest and lead V1. In this study only ECG recordings of lead II are applied to detect the heartbeat for it gives a good view of the P wave, is most commonly used to record the rhythm strip [24]. The other dataset is China Physiological Signal Challenge (CPSC) dataset [25] containing 9831 12-lead ECG recordings obtained from 9458 subjects sampled at 500 Hz. The subject-specific dataset is obtained by an ECG sensor attached on the chest of 80 persons. Each person collected ECG signals of 1 min for 5 times to get five-minute long record in comply with AAMI standards [26].

The public ECG signals used in the paper is preprocessed as shown in Figure 3. Before feeding into the neural network, these raw signals should be preprocessed to retrieve heartbeats (denoting a cardiac cycle) of a fixed length. Then the heartbeats are shuffled and rearranged to construct different datasets.

4.1 | Signal preprocessing

ECG signals from different source need to be preprocessed to obtain the samples of same length and resolution to be fed into the CNN model. The signals from CPSC are downsampled from 500 to 360 Hz to comply with MIT-BIH records. All the ECG signals are denoised and filtered to remove baseline wander using Daubechies wavelet [27]. Each ECG signal is normalized with Z-score normalization in the range of (0,1), that is, to achieve standard deviation of 1 and zero mean. Further, the ECG signals are segmented according to the location of R peak using Pan-Tompkins algorithm [28], which is regarded as the identification of a cardiac cycle. The length of each segment is fixed to 600 ms (200 ms before the R peak and 400 ms after) with 216 sample points. Finally, each segment is labelled according to the annotations provided by the public database.

4.2 | Datasets construction

The quality of the proposed model is directly related to the size and quality of training data. In many datasets publicly available, the number of records of different categories is highly unbalanced. For MITDB, the number of ECG recordings labelled with NSR is much more than that of ECG beats labelled with other categories. While for CPSC, it contains 207 records of LBBB while 1695 records of RBBB. A neural model will take the risk of performance degradation with training dataset of biased distribution, for the model would favour the dominating classes. To solve this problem, we need to balance the number of records in each category.

In this study, two different datasets A and B are constructed for performance evaluation. Dataset A includes full and unbalanced ECG data with the heartbeats retrieved only from MITDB as shown in Table 1. Dataset A is divided into training set, validation set and testing set at a ratio of 70%:10%:20% randomly.

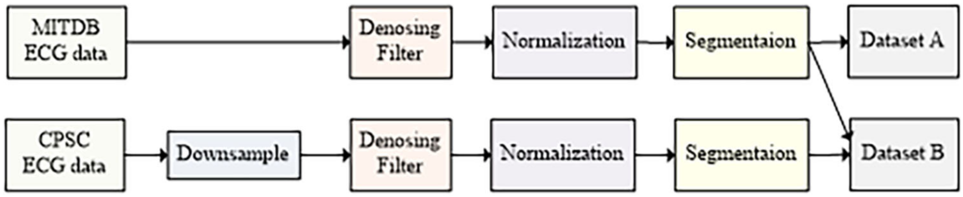

FIGURE 3 The preparation of datasets.

TABLE 1 Distribution of heartbeats in dataset A.

Arrhythmia types	Number of heartbeats
Normal sinus rhythm (NSR)	75,020
Left bundle branch block (LBBB)	8072
Right bundle branch block (RBBB)	7255
Atrial premature beats (APB)	2546
Premature ventricular contraction (PVC)	7129
Total number	100,022

TABLE 2 Distribution of heartbeats in dataset B.

Arrhythmia types	Training set and validation set	Test set	Total
NSR	16,000	15,004	31,004
LBBB	16,000	1614	17,614
RBBB	16,000	1451	17,451
APB	16,000	509	16,509
PVC	16,000	1425	17,425
Total number	80,000	20,003	100,003

Dataset B is a balanced dataset with records from MITDB and CPSC. We complement the number of the minority classes of MITDB using records of CPSC and thus make each category contains equal numbers of heartbeats (16,000). To conduct the performance comparation in a fair and sound way, dataset A and B all contain a total of approximately 100,000 ECG heartbeats in summary. Dataset B is divided into a 70% training set, 10% validation set and 20% testing set similar with dataset A as denoted by Table 2. The training set and validation set include a total of 80,000 segments. The test set contains 20,000 segments randomly selected from MITDB and CPSC without any augment. The strategy is to make sure the performance of the model is evaluated with real data sample.

5 | EXPERIMENTS AND RESULTS

The proposed model is developed and trained using Python with TensorFlow library [29]. The experiments were performed on a computer with 1 Intel Core i9-9900K CPU at 3.6 GHz, NVIDIA Quadro RTX5000 and 64GB memory. For training the model, the categorical-cross-entropy loss function was used [30]. Adam optimization method [31] was used for optimizing the model with learning rate as 0.001, beta1 = 0.9, and beta2 = 0.999. The procedure was repeated ten times to complete the tenfold training and validation plus test.

5.1 | Performance metrics

The performance of the proposed model is evaluated with the following statistical measures as shown in Equations (8)–(11): Sensitivity (Sen), Specificity (Spe), Precision (Pre), and Accuracy (Acc). Sen measures the ability of the model not to miss abnormal heartbeat, and Spec evaluates how well our model does not misjudge normal heartbeat. Pre measures the correctly predicted positive observations. Acc represents the overall performance of the model in properly classifying heartbeat. TP (True Positive) and TN (True Negative) indicate the number of heartbeats correctly predicted, while FP (False Positive) and FN (False Negative) indicate the number of heartbeats not predicted as labelled.

$$Sen = \frac{\#TP}{\#TP + \#FN} \tag{8}$$

$$Spe = \frac{\#TN}{\#TN + \#FP} \tag{9}$$

$$Pre = \frac{\#TP}{\#TP + \#FP} \tag{10}$$

$$Acc = \frac{\#TP + \#TN}{\#\,(TP + TN + FP + FN)} \tag{11}$$

For each class x, the F1 score is denoted as $F_{1x}$and computed as Equation (12), and the average F1 score of the model is evaluated as Equation (13).

$$F_{1x} = \frac{2\,(Sen * Pre)}{Sen + Pre} \tag{12}$$

$$F_1 = \frac{1}{5}\,(F_{11} + F_{12} + F_{13} + F_{14} + F_{15}) \tag{13}$$

5.2 | Experiment result

Firstly, we investigate the impact of dataset on the model. Dataset A is full and unbalanced, while dataset B is balanced.

TABLE 3 The performance of model on datasets A.

	Sen	Spe	Pre	Acc	F_{1x}
NSR	0.9570	0.8214	0.9000	0.9064	0.9276
LBBB	0.9282	0.7607	0.8270	0.8532	0.8747
RBBB	0.9416	0.6932	0.8230	0.8429	0.8783
APB	0.8596	0.6875	0.8571	0.8055	0.8584
PVC	0.9032	0.6768	0.8400	0.8246	0.8705
F1	–	–	–	–	0.8819

TABLE 4 The performance of model on datasets B.

	Sen	Spe	Pre	Acc	F_{1x}
NSR	0.9383	0.8227	0.9050	0.8970	0.9214
LBBB	0.9416	0.8128	0.8710	0.8866	0.9049
RBBB	0.9352	0.7647	0.8800	0.8753	0.9067
APB	0.8892	0.8041	0.9171	0.8644	0.9030
PVC	0.9617	0.7647	0.8800	0.8912	0.9191
F1	–	–	–	–	0.9110

TABLE 5 The F1 score of different network architecture.

	NSR	LBBB	RBBB	APB	PVC	Average F1
CNN	0.9137	0.8541	0.8812	0.8236	0.8566	0.8658
CNN-RNN	0.8943	0.8766	0.8711	0.8690	0.8752	0.8772
CNN-RNN-attention	0.9213	0.9049	0.9067	0.9030	0.9051	0.9082

Denoted by Tables 3 and 4, the average F1-scores of dataset A and B are 0.8819 and 0.9110 respectively. There is an increase in average F1-score for dataset B. The improvement is achieved by increase of F1 scores of the four arrhythmias accompanied by the slight decrease of that of NSR, but it is acceptable in that all the 5 classes achieve the same classification performance. It can be observed that the model training on dataset B achieved better performance as compared to dataset A, which is caused by low numbers of the four arrhythmias used in dataset A for training.

Then we investigate the performance comparison between the model of different architecture with subject-specific dataset. We compared the performance measures of proposed model with two different models. The first model (denoted as CNN in Table 5) used the used the structure described in Figure 2a with a full connection layer and a classification layer. The second model (de-noted as CNN-RNN) is our model without applying the attention mechanism. The F1 scores of the three models are shown in Table 5. Based on Table 5, it can be noted that model with RNN layer yielded better performance as compared to it without RNN layer. This is because the inclusion of the GRU layer captures more variations in the large number of ECG signals during training and hence helped to achieve better results. Attention mechanism highlights noteworthy features on a global scale and help to discriminate between beats of different classes, and yielding an F1 score of 0. 9082.

6 | DISCUSSION

The focus of this paper comes from several issues arose in the cardiac arrhythmia classification task: the architecture of neural network, the variability between patients, and the imbalance of dataset will impact on the ECG signal classification.

ECG signal is typical biomedical time series. The architecture of neural network for time series classification is the focus of researchers for a long time. In recent years, CNN model has proved superior classification accuracy for its feature extraction capability. Cardiac arrhythmia classification method based on CNN divides the ECG records into short segments of several seconds and outputs classification result every segment. Therefore, the method may not make full use of the entire record. Different from the separate classification of single ECG segment, we observed that adding an RNN layer can take beat-to-beat variation into account and thus model time-dependent progress of disease more precisely.

When the architecture of classification model is determined, we investigated the impact of dataset for training the model. For training the general model, one of the main limitations is that all public available datasets come from the real-world hospital and the records of classes are highly imbalanced. Generally, the record number of normal sinus rhythm is far more than other categories. Machine learning methods are often difficult to learn when a class is dominant. Using the imbalanced dataset to train the model, the classes with less records in the dataset will obtain lower positive predictive value. To generate reliable prediction results for the new data, the training dataset should be evenly distributed with nearly equal number of records in each class.

Another issue improving the unreliability of classification result is the patient variability. General model ignores the difference between patients and always suffer from performance degradation when applying to testing set with different data distribution. Many researches apply transfer learning mechanisms to overcome the shift of data distribution. In our method, we conducted global training and subject-specific training. Firstly, a general model without consideration of the variations between patients is trained based on common ECG database. Then, using a few minutes of ECG signals from one specific patient, the model is fine-tuned to find the personal characteristic pattern. Finally, the model is used to classify the ECG signals from the specific patient perpetually.

7 | CONCLUSIONS

In this study, an automatic classification model for cardiac arrhythmia combing CNN and RNN is proposed. The main contributions of this work can be summarized in three aspects:

1. Combining the benefits of both CNN and RNN architectures is helpful for the early diagnosis of cardiac arrhythmia. The subtle abnormal signs can be detected by CNN in the ECG signal before the disease have risky impact on the patient, and RNN can monitor the signs developing over time. Long-term dependency among the feature sequence is important to in that the sequence contains more information than separated heartbeat. RNN can help identify the possible abnormalities in the long-term ECG signals with the augment of attention mechanism, which will highlight pathological episodes and have a great potential to support clinicians for cardiac diagnosis in long-term ECGs.
2. To overcome the high variability between patients, the training of the model is conducted by two steps: global training and subject-specific training. The global training uses public dataset to get a generalized model. The subject-specific training uses patient-specific dataset to fine-tune the generalized model and get a subject-specific model. The experiment result shows the model achieved a satisfactory performance on the personal dataset.
3. We adopt a variety of health data sources which are shuffled and reallocated to avoid the class imbalance problem. We use a balanced training set and evaluate the proposed model with actual test set in which the availability of abnormal class is limited to show the generalization of the model. The results demonstrate that the proposed model method can provide a robust solution to the class imbalance problem in medical data.

AUTHOR CONTRIBUTIONS

Jie Sun: Conceptualization; Data curation; Funding acquisition; Methodology; Writing - original draft; Writing - review & editing.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in MIT-BIH Arrhythmia Database at https://doi.org/10.1161/01.CIR.101.23.e215, reference number [18], and China Physiological Signal Challenge dataset at https://doi.org/10.1166/jmihi.2018.2442, reference number [19]. Personal-specific research data are not shared.

ORCID

Jie Sun https://orcid.org/0000-0003-2996-7613

REFERENCES

1. National Center for Cardiovascular Diseases, China: Report on cardiovascular diseases in China 2018. Encyclopedia of China Publishing House, Beijing (2019)
2. Pinto, J.R., Cardoso, J.S., Lourenço, A.: Evolution, current challenges, and future possibilities in ECG biometrics. IEEE Access 6, 34746–34776 (2018)
3. Antoni, B.L.: Clinical Electrocardiography: A Textbook. 4th ed. Wiley, Hoboken, NJ (2012)
4. Faust, O., Hagiwara, Y., Hong, T.J., Lih, O.S., Acharya, U.R.: Deep learning for healthcare applications based on physiological signals: A review. Comput. Methods Programs Biomed. 161, 1–13 (2018)
5. Lyon, A., Mincholé, A., Martínez, J.P., Laguna, P., Rodriguez, B.: Computational techniques for ECG analysis and interpretation in light of their contribution to medical advances. J. R. Soc. Interface 15(138), 1–18 (2018)
6. Hannun, A.Y., Rajpurkar, P., Haghpanahi, M., Tison, G.H., Bourn, C., Turakhia, M.P., Ng, A.Y.: Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. Nat. Med. 25(1), 65–69 (2019)
7. Baloglu, U.B., Talo, M., Yildirim, Ö., Tan, R.S., Acharya, U.R.: Classification of myocardial infarction with multi-lead ECG signals and deep CNN. Pattern Recognit. Lett. 122, 23–30 (2019)
8. Acharya, U.R., Fujita, H., Oh, S.L., Hagiwara, Y., Tan, J.H., Adam, M., Tan, R.S.: Deep convolutional neural network for the automated diagnosis of congestive heart failure using ECG signals. Appl. Intell. 49, 16–27 (2018)
9. Maknickas, V., Maknickas, A.: Atrial fibrillation classification using QRS complex features and LSTM. Proc. CinC 2017, 1–4 (2017)
10. Chang, Y., Wu, S., Tseng, L., Chao, H., Ko, C.: AF detection by exploiting the spectral and temporal characteristics of ECG signals with the LSTM model. Proc. CinC 45, 1–4 (2018)
11. Saadatnejad, S., Oveisi, M., Hashemi, M.: LSTM-based ECG classification for continuous monitoring on personal wearable devices. IEEE J. Biomed. Health Inf. 24(2), 515–523 (2020)
12. Sawant, N.K., Shivnarayan, P.: Diagnosis of cardiac abnormalities applying scattering transform and Fourier-bessel expansion on ECG signals. Proc. CinC 48, 1–4 (2021)
13. Yildirim, Ö., Baloglu, U.B., Tan, R.S., Ciaccio, E.J., Acharya, U.R.: A new approach for arrhythmia classification using deep coded features and LSTM networks. Comput. Methods Programs Biomed. 176, 121–133 (2019)
14. Lui, H.W., Chow, K.L.: Multiclass classification of myocardial infarction with convolutional and recurrent neural networks for portable ECG devices. Inf. Med. Unlocked 13, 26–33 (2018)
15. Swapna, G., Soman, K.P., Vinayakumar, R.: Automated detection of diabetes using CNN and CNN-LSTM network and heart rate signals. Procedia Comput. Sci. 132, 1253–1262 (2018)
16. Murugesan, B., Ravichandran, V., Ram, K., Preejith, S.P., Joseph, J., Shankaranarayana, S.M., Sivaprakasam, M.: ECGNet: Deep network for arrhythmia classification. In: 2018 IEEE International Symposium on Medical Measurements and Applications (MeMeA). pp. 1–6, Rome, Italy (2018)
17. Hu, D.: An introductory survey on attention mechanisms in NLP problems. In: Advances in Intelligent Systems and Computing. p. 1038. Springer, Cham (2019)
18. Guo, M.H., Xu, T.X., Liu, J.J., et al.: Attention mechanisms in computer vision: A survey. Comput. Visual Media 8, 331–368 (2022)
19. Sigurthorsdottir, H., Van, Z.J., Delgado-Gonzalo, R., Lemay, M.: ECG classification with a convolutional recurrent neural network. In: 2020 Computing in Cardiology. Rimini, Italy, pp. 1–4.(2019)
20. Qiao, L., Cao, X., Ma, T., Li, H., Sun, J.M.: Multilevel knowledge-guided attention for modeling electrocardiography signals. In: IJCAI 2019. Macao, China, pp. 5888–5894 (2019)
21. Hu, J., Shen, L., Albanie, S., Sun, G., Wu, E.: Squeeze-and-excitation networks. IEEE Trans. Pattern Anal. Mach. Intell. 42(8), 2011–2023 (2020)
22. Ioffe, S., Szegedy, C.: Batch normalization: Accelerating deep network training by reducing internal covariate shift. ICML 37, 448–456 (2015)
23. Goldberger, A.L., Amaral, L.A.N., Glass, L., Hausdorff, J.M., Ivanov, P.C., Mark, R.G., Mietus, J.E., Moody, G.B., Peng, C.-K., Stanley, H.E.: PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. Circulation 101(23), e215–e220 (2000)
24. Francis, M., William, J.B., John, C.: ABC of Clinical Electrocardiography. 2nd ed. p. 3. Blackwell Publishing Ltd., Massachusetts, USA (2008)

25. Liu, F., Liu, C.Y., Zha, L., Zhang, X., Wu, X., Xu, X., Liu, Y., Ma, C., Wei, S., He, Z., Li, J., Ng, E.: An open access database for evaluating the algorithms of electrocardiogram rhythm and morphology abnormality detection. J. Med. 8, 1368–1373 (2018)
26. AAMI.: Recommended practice for testing and reporting performance results of ventricular arrhythmia detection algorithms. AAMI Standards and Recommended Practices, 2.2, 921–944 (1987)
27. Singh, B.N., Tiwari, A.K.: Optimal selection of wavelet basis function applied to ECG signal denoising. IEEE Trans. Circuits Syst. 16(3), 275–287 (2006)
28. Pan, J., Tompkins, W.J.: A real-time QRS detection algorithm. IEEE Trans. Biomed. Eng. BME-32(3), 230–236 (1985)
29. Abadi, M., Barham, P., Chen, J., et al.: TensorFlow: A system for large-scale machine learning. In: USENIX Symp Oper Syst Des Implement (OSDI). GA, USA, pp. 265–283 (2016)
30. Deng, L.Y.: The cross-entropy method: A unified approach to combinatorial optimization, Monte-Carlo simulation, and machine learning. Technometrics 48, 147–148 (2006)
31. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. In: ICLR 2015. CA, USA (2015)

How to cite this article: Sun, J.: Automatic cardiac arrhythmias classification using CNN and attention-based RNN network. Healthc. Technol. Lett. 10, 53–61 (2023). https://doi.org/10.1049/htl2.12045